

REVIEW OF *MULTIPLE AFFORDANCES OF LANGUAGE CORPORA FOR DATA-DRIVEN LEARNING*

Multiple Affordances of Language Corpora for Data-driven Learning

Agnieszka Leńko-Szymańska & Alex Boulton

2015

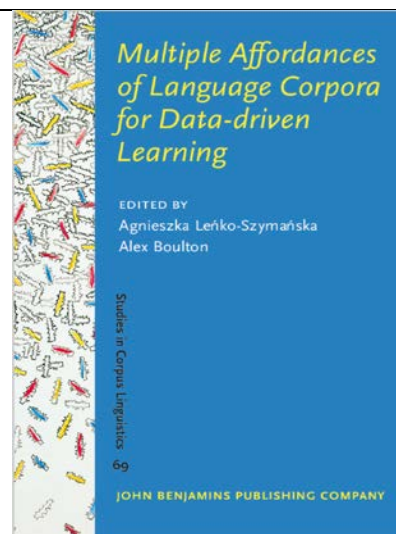
ISBN: 978-90-272-0377-9

US \$149.00

312 pp.

John Benjamins Publishing Company

Amsterdam



Review by Hyung-Jo Yoon and Jungmin Lim, Michigan State University

Since the first availability of corpora for language analysis in the 1960s, the technologies related to corpus compilation, annotation, and analysis have greatly improved. Now many corpus researchers are showing increasing interest in how to use corpus data as a useful pedagogical tool. Contributing to this line of interest, *Multiple Affordances of Language Corpora for Data-driven Learning* presents a range of pedagogical applications of corpora. This volume, composed of 12 chapters, addresses three major themes: corpora for language learning (Part I), corpora for skills development (Part II), and corpora for translation training (Part III). All chapters emphasize the varied affordances of corpora that enable learners to have greater exposure to authentic language in context. Interestingly, the pedagogical use of corpora introduced in this volume is applied to a wide range of language skills for second language learners as well as to language specialists, such as translators.

The introduction by the editors, Leńko-Szymańska and Boulton, starts with the notion of the direct use of corpora and their multiple affordances for language teaching. Recognizing the currently marginal role of direct corpus use in class settings, the authors suggest a range of advantages of using corpus data: access to authentic language, increased language awareness, and learner autonomy. Further, they provide effective justifications for inductive data-driven learning (DDL) activities and solutions to some obstacles such as extra demands and its predominant focus on grammatical structures rather than meaning.

In the second chapter, to address what she views as a lack of theoretical underpinnings in many corpus studies, Flowerdew links corpus applications to language learning theories. Most notably, she explains the relevance of the *noticing hypothesis* and how it relates to the *inductive learning principles* of DDL. She further discusses the collaborative dimensions of DDL activities, such as scaffolding (in relation to Vygotskian sociocultural theories), that promote learner autonomy and interaction.

In Chapter 3, Tribble describes how corpus applications have evolved and will further develop by reflecting on his own hands-on experience of corpus applications. He reviews the literature on direct corpus applications in classroom settings. Furthermore, this chapter focuses on a summary of a survey on teachers' corpus tool preferences and perceived benefits to the students, suggesting that many teachers have positive attitudes toward corpus use, despite the existence of notable difficulties. These introductory chapters clearly help readers better understand the history and development of in-class corpus

applications.

Part I consists of three chapters (Chapters 4, 5, and 6) that commonly illustrate alternative corpus-based methods that focus on chunks. The authors of these chapters discuss how language learners benefit from new methods and tools aimed at facilitating autonomous corpus searches. In Chapter 4, Aston proposes a methodology which uses speech corpora and aligned transcripts to teach phraseological items. In the first part of this chapter, he focuses on the necessity of using speech corpora and corresponding transcripts for a better understanding of distinctive prosodic features. Then, with some exemplary *n*-gram analyses (i.e., presenting high-frequency contiguous sequence of *n* number of words) of a speech corpus (a collection of TED talk videos), Aston suggests how teachers can assist their students in identifying and interpreting high-frequency *n*-grams more effectively with the help of speech corpora. He further discusses student willingness to learn phraseological items as a prerequisite for successful implementation of this new method. The fifth chapter by Thomas also emphasizes the importance of multi-word units (here collocations) to avoid deviations. Using Sketch Engine, the author demonstrates a specific procedure named Collocation Plus, in which learners are encouraged to explore lexicogrammatical features surrounding target collocations and then to produce sentences using the collocations to which they have been exposed during the previous stages. The author predicts that this procedure promotes discourse-level understanding of vocabulary in relation to its specific corresponding construction.

In Chapter 6, to address the issue of low accessibility of corpus resources to beginning EFL students, Chujo, Oghigian, and Akasegawa introduce a new DDL tool (GPPS) appropriate for low-proficiency students. Distinguished from other concordance tools that present search results in authentic context (i.e., KWIC), the GPPS is adjustable for different proficiency levels and shows sentence-level search results with translations in a separate column. Unlike the previous two chapters advocating full text use, the authors argue for the exploitation of sentence-level results rather than fully contextualized ones from the GPPS so that low-proficiency students can better develop their familiarity with sentence structures.

Part II suggests potential uses of corpora in the classroom to build language skills and strategies, including academic writing (Chapter 7) and reading comprehension (Chapters 8 and 9). In Chapter 7, Charles demonstrates corpus-based English for Academic Purposes writing courses in which graduate students and post-doctoral students from diverse academic backgrounds build and analyze their own corpus from academic journals and share the lexical items with peers. She also illustrates how to prepare students to exploit different functions of AntConc, a freely available corpus analysis tool, as their discovery tools. Learners reported this learning method helped them to construct representations of discipline-appropriate discourse.

Turning to the reading skills, Chapter 8 by Babych examines how teaching materials using news articles in which connectors are annotated with their functional categories can enhance student awareness of text structures. In addition, this chapter addresses special population learning cognates of a previously learned language. Babych extracted connectors at the beginning of sentences in a large corpus of each language and classified them into 16 functional categories. She constructed a corpus of online news for each language in which the connectors were highlighted and glossed with their functions. After applying this corpus as online reading materials, the author detected significant gains in global comprehension, translation, and vocabulary among her study participants. She concludes the chapter with practical suggestions for developing reading skills.

Chapter 9 by Curado Fuentes reports on an experimental study into the development of reading comprehension skills, in which the effectiveness of DDL activities using news keywords was examined in comparison with traditional paper-based activities. Fully aware of the fact that textual features of newspapers are highly authentic and complex, the author investigated how highly contextualized DDL activities led to intermediate-level student comprehension of the keywords extracted from newspapers. The findings from the post-test, questionnaire, and short writing point to the benefits of DDL activities on high learner motivation and favorable attitudes.

Part III targets translator trainers and suggests corpus-based tasks for a variety of translation skills: web-based practicing (Chapter 9), raising awareness of interference of source text (Chapter 10), and highlighting genre-specific features of film subtitle translation (Chapter 11). In Chapter 9, Molés-Cases and Oster illuminate a gap between the abundance of corpus resources for translators and the lack of knowledge of these resources by trainees. The authors introduce two exemplary corpus-based tasks delivered in webquests about pronunciation and pragmatics in German. The webquests are structured in a way that helps learners develop their autonomous use of authentic language in corpora and interact with peers and teachers while sharing feedback. Based on active student participation and survey responses, the authors confirm the effectiveness of webquests as learning tools for future translators.

In terms of awareness-raising tasks for translator trainees, Marco and van Lawick introduce a successful pedagogical attempt to enhance translator trainee awareness of source text interference in Chapter 10. The authors propose using comparable corpora to help learners better notice unnatural style of translated texts due to traces of original text. Participants performed cyclical tasks of analyzing translated Catalan texts in comparison with natural Catalan texts using WordSmith Tools until they achieved a certain level of awareness on source text interference. Drawing on the revisions participants made on their writing after receiving training, this learning cycle was reported to be positive for learning and for familiarizing students with corpora for translation practices. The next chapter by Sotelo illustrates how trainee awareness on a specific subtitle genre can be promoted by DDL using multimedia corpus. She proposes four tasks for students to notice features of translated texts such as condensations, omissions, and additions. Based on her observations of student engagement with tasks and responses to their learning experience, Sotelo remarks positively on the pedagogical application of multimedia corpus on translator training and its match with technological competence.

This book ends with Boulton's proposal for practical applications of DDL. As clearly demonstrated by the chapters above, DDL is promising for language development in diverse contexts, but it usually requires learners to access an appropriate corpus and use specific corpus analysis tools; these preconditions surely demand a good amount of time and training. Boulton recommends the use of the Web as a large corpus and powerful search engines, such as Google, as DDL tools. Despite having less accuracy, results from carefully arranged search terms on the web could be sufficient for many learning contexts. The author highlights the practicality and ease of the Web, thus leading to a great benefit for teachers and learners.

As the title of this book suggests, the overarching theme of this edited volume is to introduce a variety of pedagogical affordances elicited from corpora and corpus-analysis tools. The chapters are easy to follow, and on the organizational side, one can tell that the editors put great effort into showing various aspects of DDL from theoretical and broad language learning contexts to more specialized settings of translator training. This book has its unique value in that it introduces a wide range of corpus types that differ in size, language type, and language modality. This book also provides information about how to use various corpus analysis tools to fit varying needs of target students. In addition, this volume suggests and exemplifies how language teachers can design and sequence tasks for DDL. For example, individual tasks delivered in the online platform can promote autonomous learning, while in-class group tasks can increase interactions between students. The book focuses mostly on real applications of corpus data and tools in language learning settings. Accordingly, little attention is given to technical dimensions of corpora such as how to systematically annotate and analyze collected texts. For those seeking additional, detailed explanations of how to build corpora, one might have to look elsewhere. If one is interested in having a balanced view of corpora in terms of their annotations and pedagogical use, referring to other books focusing on corpus annotations (e.g., Kübler & Zinsmeister, 2015; Lu, 2014) can be one possibility.

Most chapters emphasized that the quality of DDL depends on how teachers construct and sequence corpus-based tasks. However, the chapters lacked detailed guidelines about how to sequence DDL tasks

in a particular setting, and they did not suggest specific criteria for the decisions about which types of corpus data and analysis tools to employ. Thus, this book can be well supplemented with other books or articles related to task-based language teaching and task sequencing. Similarly, with regard to the illustration of training procedures in some empirical chapters, the authors briefly discuss how they trained learners to use corpus tools. However, as admitted by Tribble in Chapter 3, DDL has not yet achieved its firm role in language classrooms. One major stumbling block has been the additional efforts required for teachers to guide their students because DDL activities cannot elicit effective learning without thorough learner training. Furthermore, in ESL settings, the classroom may consist of students with a wide variety of backgrounds. Some may have limited experience of using computers, and thus low digital literacy, making it extremely difficult to teach them how to use corpus tools. Given the pivotal role of learner training for the successful implementation of DDL activities, it would have been great to include more detailed descriptions of how language learners with different backgrounds and learning goals can be motivated and trained for effective corpus-based text analysis.

Nonetheless, this book is a valuable resource for practitioners and researchers who hope to learn the value and effectiveness of data-driven learning and to understand how to use various types of corpora and analysis tools in educational settings. The rationale for choosing DDL over traditional learning is its possibility to expose learners to authentic texts and then to draw their attention to target forms inductively in accordance with the principles of *focus on form*. Given the consistent findings showing the necessity of an explicit focus on form in meaning-based language learning (Norris & Ortega, 2000), language teachers may greatly benefit from this book. By reading this book, teachers who have vague ideas of how to use corpus data for language teaching can shape their ideas into specific plans aimed at guiding learners in using corpora for their own learning. This volume is a valuable resource to possess. The book is also for those interested in DDL research, corpus building, and corpus tool development. Reading this book will lead them to clearly understand the true value of DDL as an effective and flexible teaching method that makes full use of current technological advances.

ABOUT THE AUTHORS

Hyung-Jo Yoon is a Ph.D. student in Second Language Studies at Michigan State University. While pursuing a degree, he has been working as an English instructor at the English Language Center. His research interests include second language writing, language assessment, and computational analysis of natural language.

E-mail: yoonyhyu9@msu.edu

Jungmin Lim is a doctoral student in the Second Language Studies program at Michigan State University. Her research interests are second language writing, computer-assisted language learning, and language assessment.

E-mail: limjung2@msu.edu

REFERENCES

- Kübler, S., & Zinsmeister, H. (2015). *Corpus linguistics and linguistically annotated corpora*. London: Bloomsbury.
- Lu, X. (2014). *Computational methods for corpus annotation and analysis*. NY: Springer.
- Norris, J. M., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning*, 50(3), 417–528.